Information Strategic Plan:

Modernizing the Safe Drinking Water Information System for the Public Water System Supervision Program

Office of Water
EPA 816-R-04-001
January 2004
www.epa.gov/safewater

VISION STATEMENT

The Office of Ground Water and Drinking Water's (OGWDW) Information Strategic Plan supports the objective of improving the quality of public health in the United States through the provision of high quality and affordable drinking water supplies. The timely, efficient and effective gathering, analysis, and access to complete, high quality and integrated data are critical to the attainment of this objective. The regulated community, states, EPA, other governmental bodies, stakeholders, and the general public are key partners in this ongoing effort.

TABLE OF CONTENTS

Appendices:

       Appendix A: EPA ENTERPRISE ARCHITECTURE CORE COMPONENTS
       Appendix B: OGWDW INFORMATION ARCHITECTURE CORE COMPONENTS

Attachments:
       Attachment 1: List of OGWDW Data Systems

**OFFICE OF GROUND WATER AND DRINKING WATER
INFORMATION STRATEGIC PLAN**

I. EXECUTIVE SUMMARY

Public demand grows for easily accessible, integrated information about drinking water. The Office of Ground Water and Drinking Water (OGWDW) faces many challenges in managing its information resources to support the U.S. Environmental Protection Agency (EPA or Agency) mission of protecting human health. OGWDW has responded and will continue to respond to the American public's demand for information. As demands increase faster than information resources do, OGWDW must take advantage of advances in information technology (IT) to optimize collection, processing, storage, and access to its data.

A. The Agency Information Management Goals:
1) **Use:** Improve the use of environmental information to support decision-making, activity cost accounting, and results-based management
2) **Data:** Collect appropriate data and provide high-quality and integrated information
3) **Technology:** Strengthen EPA 's information infrastructure to improve Agency operations and the security, collection, and exchange of information
4) **Access:** Enhance public access to useful and understandable information
5) **Governance:** Adopt an enterprise-wide approach to make and implement information management decisions
6) **People:** Invest in human capital

Using these goals, OGWDW has developed a conceptual framework for managing its information resources for the Public Water System Supervision (PWSS) Program modernization effort**.** This document describes the information management goals (through December 2004) for OGWDW.

B. OGWDW Information Management Goals through 2004:
1) Improve the **use** of data by working to improve its accuracy, timeliness and completeness
2) Continue to collect appropriate **data** and expand its collection and its integration ability to accommodate new business needs
3) Improve the security of its data flow while enhancing its ease-of-exchange through new **technology** use
4) Enhance public **access** to data through new reporting tools that use a modern data warehouse
5) Facilitate **governance** by full participation with the Agency Enterprise Architecture
6) Leverage the skills and abilities of its **people** to do more in-house and so reduce costs and increase job satisfaction

OGWDW has made strides towards achieving these goals in the PWSS Program. In the future, with continued support from senior management and state and regional partners, OGWDW will create modern data flows for each of its ground water and drinking water programs that will support public health and environmental protection and will meet the needs of stakeholders.

II. INTRODUCTION

A. This Information Strategic Plan describes OGWDW's approach for the application of technology to manage the data flows currently received from states and Regions under the Public Water System Supervision (PWSS) Program.

To enhance its decision-making for public health protection, the Office of Ground Water and Drinking Water (OGWDW) will expand and maintain its modern data warehouse for the drinking water compliance and source water data received from states and Regions, ultimately through EPA's secure central data exchange (CDX). [Note: Modernization of the data systems in no way changes reporting requirements under the Safe Drinking Water Act and its regulations. Data that has been reported in the past will continue to be reported in the future. Modernization will primarily change *how* the data is reported.]

B. OGWDW will implement the first stage, addressing the public water system (PWS) facility and compliance data, by the end of 2004.

The SDWIS-FED data warehouse is already modernized to extract, receive, transmit and store SDWIS and Unregulated Contaminant data (a small subset of all PWS data, used for analyzing emerging drinking water contaminants). When the warehouse is fully operational, EPA will have a comprehensive data set in one place. In the Safe Drinking Water Information System (SDWIS) data warehouse of the future, EPA will manage its ground water, drinking water, and source water protection data responsibilities under the Safe Drinking Water Act (SDWA). EPA will make these data publicly available as extensively as possible. The office will implement the second stage of this plan, including data flows for Underground Injection Control (UIC) compliance and Source Water Protection (SWP) in the data warehouse, after 2004.

C. This Information Strategic Plan document addresses modernizing the two existing Safe Drinking Water Information Systems supporting the data flow for the PWSS Program, SDWIS-FEDERAL and SDWIS-STATE.

Appendix A and Appendix B describe the Agency's and OGWDW's information architecture plans, including descriptions of the UIC and SWP programs. A list of the current and/or planned data systems in OGWDW is included as an attachment. This document and appendices also provide details on Agency responsibilities, the business questions and decisions associated with these responsibilities, and the data needed to support these questions and decisions.

III. MANY DATA FLOWS, ONE WAREHOUSE

A. The OGWDW Information Strategic Plan must be a "living" plan.

Many driving forces for change influence the application of information technology.
- L**egislative drivers**, such as the Safe Drinking Water Act, the Clean Water Act, the Government Performance and Results Act, the Clinger-Cohen Act (manage risks in IT acquisitions), the Government Paperwork Elimination Act , and the Government Information and Security Reform Act.
- **Business drivers**, such as increasing expectations for e-government; need for secure, high-quality, meaningful data; increased emphasis on accountability and performance measures (including integration with the Agency's Strategic Plan); and, increasing integration of environmental information.
- **Technology drivers**, such as wide acceptance of the personal computer and the Internet; rapid evolution of computing and networking technologies at lower cost; and, development of object oriented computer programming languages that allow building larger, more complex, more reliable applications at lower costs with reduced maintenance.

As OGWDW modernizes its information systems, a "living" Information Strategic Plan is critical to provide a logical course of action for accommodating these drivers, to document decisions made, and to communicate those decisions to stakeholders

B. A unifying concept of the plan is to maintain one data warehouse for all data flows that support the nation's ground water and drinking water program.

Under the Safe Drinking Water Act, EPA administers three programs: the Public Water Supply Supervision Program (PWSS), the Underground Injection Control (UIC) Program, and the Source Water Protection (SWP) Program. Of these three programs, the most significant data flow supports the PWSS Program, which involves approximately 167 thousand public water systems.
This document will focus on the PWSS Program.

1. Description of the PWSS Program:

Essentially, SDWA allows the states to assume primary enforcement responsibility (primacy) for the rules and regulations developed by EPA. Today, there are 57 drinking water primacy agencies: 49 states, EPA Region 8 (for Wyoming), the Navajo Tribe, District of Columbia, US Virgin Islands, Northern Marianas, American Samoa, Guam, and Puerto Rico. EPA retains primacy in nine of its regional offices for Native American tribes

The PWSS Program is somewhat mature, having been created in the 1970's and modified and updated through several legislative amendments over the years. However, the basic tenets of the roles and responsibilities of EPA and the states have been stable.

EPA maintains principal responsibility for identifying contaminants which have adverse impacts on human health that are likely to occur in drinking water supplies and for developing health based criteria and regulations for control of the contaminants. In the event a state fails to enforce these regulations, EPA is authorized to intervene and take enforcement action to restore compliance. In addition, EPA ensures that states implement the rules in accordance with regulatory requirements. Further, SDWA authorizes EPA to assist in funding the states' costs of program implementation by providing annual grants commensurate with their drinking water program workloads

2. Data Needs of the PWSS Program:

Four major areas of the PWSS Program supported by the SDWIS data flows are:
- Rule management (including rule determination, rule development, rule implementation and rule review);
- Oversight of state programs (including the drinking water implementation and enforcement programs);
- Technical and compliance assistance; and,
- Outreach.

SDWIS provides access to specific sets of data, in many cases on a routine basis, to carry out its responsibilities in these four areas. In general, SDWIS data characterizing the PWSs (e.g., owner type, population served, source type, geographic data, etc.) is needed for rule management to assist in identification of controls and alternative analyses, as well as for grants decisions, outreach/assistance and state program oversight. SDWIS data on violations and enforcement actions are used for state program and enforcement oversight, as well as EPA compliance support activities and outreach/assistance activities.

For the PWSS Program alone, the current OGWDW architecture must receive, store, and access data for 90 contaminants at approximately 167 thousand water systems with data collection frequencies ranging from every several hours to annually and longer. These information systems primarily rely on unsupported computer languages and software.

Historically, it has taken three separate components to manage these data:
     1) SDWIS-STATE software for state data and program management;
     2) DTF – the data transfer format for states to submit data to EPA about compliance of
         water systems; and,
     3) SDWIS-FEDERAL (SDWIS-FED) – the federal repository for state data on water
         system compliance.

One unified, modernized system, developed in concert with the Agency's efforts to modernize its information systems, will promote concrete improvements in public health, by enhancing EPA's ability to pinpoint public water systems for compliance assistance, or, alternatively, enforcement action.

IV. COORDINATION WITH EPA'S ENTERPRISE ARCHITECTURE DEVELOPMENT

A. OGWDW's strategic plan for modernizing its drinking water information systems must accommodate the evolving requirements of the Agency Enterprise Architecture and document the principles used to make those accommodations.

EPA is in the midst of modernizing its systems to facilitate electronic data exchange and integration by developing Enterprise Architecture (EA). The Federal Chief Information Council defines Enterprise Architecture as follows:

> *"Enterprise Architecture – a strategic information asset base, which defines the mission, the information necessary to perform the mission, the technologies necessary to perform the mission, and the transitional processes for implementing new technologies in response to the changing mission needs. An enterprise architecture includes a baseline architecture, target architecture, and a sequencing plan."*

The Enterprise Architecture process and principles that govern the implementation of the architecture affect its development, maintenance, and use. For implementation, principles establish the first tenets and related decision-making guidance for designing and developing information systems. Each and every phase of the systems life cycle is supported by the actions these principles require. Capital Planning and Investment Control actions are governed by the implications within the principles.

B. The OGWDW Information Strategic Plan incorporates the following ten Enterprise Architecture principles established by the Federal Chief Information Officer (CIO) Council:

- Architecture is scoped, planned, and defined based on the intended use of the architecture.
- Architecture is compliant with the law as expressed in legislative mandates, executive orders, federal regulations, and other federal guidelines.
- Architecture facilitates change.
- Architecture reflects the OGWDW and Agency's strategic plan.
- Architecture continuously changes and requires transition.
- Target architecture projects no more than three to five years into the future.
- Architecture provides standardized business processes and common operating environments (COEs).
- Architecture products are only as good as the data collected from subject matter experts and domain owners.
- Architecture minimizes the burden of data collection, streamlines data storage, and enhances data access.
- Target architecture is used to controls the growth of technical diversity.

C. This Plan incorporates these ten Agency EA principles to carry out OGWDW'S six information management goals:
- Improve use of environmental and public health data.
- Collect appropriate data.
- Strengthen information management infrastructure.
- Enhance data access.
- Adopt an Agency-wide approach to information management.
- Invest in human capital.

D. OGWDW also uses the following principles to carry out its goals:
- Keep processes and systems as simple as possible.
- Base systems development on clearly defined business needs.
- Contain information systems costs.
- Minimize reporting requirements for PWS and states.
- Maintain support for core components/processes including SDWIS-STATE and new rules.
- Employ sound data management practices and data standards.
- Improve data quality through documented procedures.
- Ensure tracking and reporting of Government Performance Results Act measures.
- Modernize in stages consistent with available funding.
- Employ proven technologies and commercial off-the-shelf software wherever feasible.

E. OGWDW participates in formulation of, and compliance with, Agency policies, guidelines and directives.

Key components of the current Agency thinking in this regard include the Exchange Network, Central Data Exchange, Facility Registry System, and Chemical and Biological Registries, data stewardship and Trading Partner Agreements. (See Appendix A for details on these components.)

F. OGWDW faces continuing challenges.

As stated earlier, the Information Strategic Plan is a "living" statement of action. Further development will consider how the drinking water information management system will incorporate evolving Enterprise Architecture components and considerations. It will also consider constraints and assumptions of the Office of Water and/or OGWDW, such as coordination/cooperation with other key Agencies and Departments, such as the Department of Homeland Security, Food and Drug Administration, United States Geological Survey, state health departments, and tribal partners. Relationships with PCS, STORET and other major water systems and the programs/business needs they support will also be addressed with emphasis on alignment and integration, where feasible.

IV. OGWDW INFORMATION STRATEGIC PLAN APPROACH

For the purposes of this document, the Ground Water and Drinking Water Program adopts the definitions, guidance and approach of the Federal CIO Council. This approach integrates the OGWDW modernization effort into the larger Office of Water and Agency Enterprise Architecture efforts.

The Federal CIO Council defines Baseline Architecture as – *"the set of products that portray the existing enterprise, the current business practices, and technical infrastructure, commonly referred to as the "As-Is" architecture."* This section describes currently existing information architecture.

The Information Strategic Plan follows the Federal CIO Council guidance format:
- **Baseline Architecture** (What exists/are doing now);
- **Target Architecture** (What is wanted/will be doing); and,
- **Transition Plan** (How to get from Baseline to Target). For each of these components, four fundamental "building blocks" of the architecture (Business Processes, Data Model(s), Applications, and Technology) are included. (Further details of the Baseline and Target Architectures for SDWIS are included in Appendix B.)


A. Baseline Architecture – Business Processes

Business Processes – OGWDW's Public Water System Supervision Program:
- Authorizes the regulation of the facilities that treat, store and distribute drinking water to taps;
- Implements the National Primary Drinking Water Regulations (NPDWR) for 90 contaminants developed and issued by EPA; and
- Implements programs to enhance water system operation.

Public Water System Supervision Program

Over 167,000 Public Water Systems Nationwide

54,064

93,210

20,559

CWSs   NTNCWSs   TNCWSs
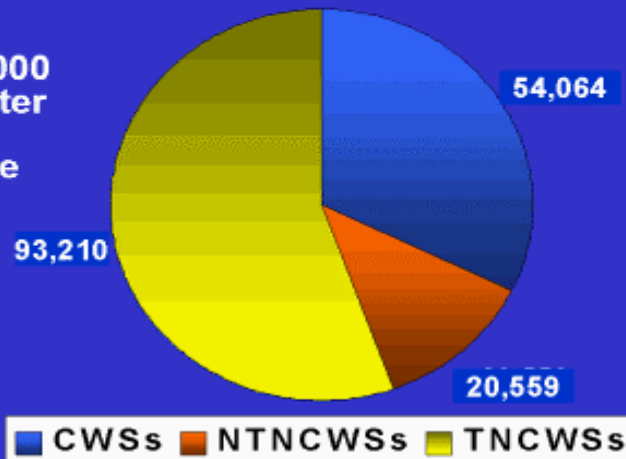
Background of Business Processes

There are currently approximately 167 thousand water systems regulated by the federal government in the U.S. Risks to different types of public water systems vary, so they are divided into three types: community water systems (CWS), transient non-community water systems (TNCWS), and non-transient, non-community water systems (NTNCWS).

Community water systems serve the majority of the population. They can be vast, serving millions of people (such as New York City or Boston) or small, serving a trailer park with 25 residents.

As shown above, the majority of PWSs are transient non-community water systems, which include roadside stops, commercial campgrounds, hotels, restaurants, and other facilities that have their own water supplies and serve a transient population at least 60 days per year. While these systems are numerous, they do not serve the majority of the population because each system only serves a small number of people. However, almost everyone is served by transient non-community water systems at some point.

To protect the health of populations served by all public water systems, National Primary Drinking Water Standards establish either the maximum concentration of pollutants allowed in, or the minimum treatment required for, water that is delivered to customers. The scientific basis for each standard is a Maximum Contaminant Level Goal (MCLG), the maximum level of a contaminant in drinking water at which no known or anticipated adverse health effects would occur.

Most standards consist of an enforceable MCL, the maximum permissible level of a contaminant in water that can be delivered to any user of a public water system. It is set as close to an MCLG as possible, taking into account the costs, benefits and feasible technologies.

For some contaminants, no reliable, economically and technologically feasible method exists to measure the contaminant, particularly at low concentrations. In these cases, EPA establishes a treatment technique (TT). A treatment technique is an enforceable procedure or level of technological performance that public water systems follow to ensure control of a contaminant. An example of a treatment technique involves protection of consumers from certain pathogens. Reliably measuring the concentration of pathogens can be cost-prohibitive. EPA has found that proper operation of filters at a certain level of performance would reliably remove the pathogens from the water. EPA has implemented regulations requiring filtration at a specified level of performance in place of setting allowable contaminant concentration standards.

In the regulatory scheme provided by SDWA, EPA conducts and/or analyzes research and other data on the public health impacts of a contaminant; evaluates treatment and control technologies and associated costs; conducts risk assessments associated with various levels of a contaminant in drinking water; and establishes an MCL or treatment technique it determines as economically achievable. EPA also establishes monitoring requirements for contaminants, specifying the number and types of samples to be collected, the frequency and locations of sampling in water systems are required to provide necessary treatment or controls, conduct necessary monitoring, an submit monitoring results to the primacy agency. The primacy agency (usually a state or tribe) has primary responsibility for determining compliance and taking necessary enforcement actions.

EPA regional offices oversee and track primacy agency/state enforcement efforts and directly enforce the regulations in Direct Implementation states. Oversight and enforcement focus on actions against significant non-compliers (SNCs). Significant noncompliance presents a potentially serious public health concern over an extended period of time (as opposed to a single monitoring violation, for example

The primacy agency submits certain information and data on PWSs and violations of regulatory requirements to EPA on a routine basis. EPA compiles this data, does quality control checks, analyzes the data, calculates SNCs, where necessary provides compliance assistance to the primacy agency or PWS, or in some instances takes federal enforcement action to compel compliance. EPA also makes data available to the public, develops national trends and statistics, prepares formal reports to Congress and uses the data to assist in further policy or regulatory development. SDWIS-FED is the EPA information management system that supports this high priority business process.

Some states have no automated information management capability or are challenged by re-engineering their own software for every new or changing drinking water standard. EPA has developed SDWIS-STATE, an information management software system designed to assist states with limited information management systems or capability of their own. SDWIS-STATE is of much broader scope and much larger than SDWIS-FED because it is designed to help the

states manage the entire PWSS Program, including additional state program requirements, such as monitoring schedules, monitoring results and comparison of results to standards. SDWIS-STATE software is also designed to extract and properly format the subset of data that states report to EPA. Information contained in SDWIS-FED is a small subset of information contained in SDWIS-STATE. Currently, 25 states, six EPA regions and two U.S. Territories are using SDWIS-STATE.

A. Baseline Architecture – *Business Processes*, Data Model(s)

Data Model(s) – Public Water System Supervision Program data are:
- Managed in SDWIS-FED, a data system designed to support OGWDW in monitoring compliance with the Safe Drinking Water Act by receiving, storing, and processing data from states;
- Managed in SDWIS-STATE, designed to assist states in managing their entire drinking water program on a day-to-day operational level; and,
- Housed in the OGWDW data warehouse.

Background of Data Models:

1. The SDWIS-FED system processes the following major categories of information reported by all PWSS programs:

- Inventory
  - Characteristics of Public Water Supply Systems (includes administrative contact information, activity status, PWS type, population served, primary source type, and owner type).
  - Water system facility and treatment data (includes flow data between sources of water through the treatment plant).
  - Locational and geographic data to support geospatial applications and source water assessments.
- Violations of the national Primary Drinking Water Standards and other implementing regulations of the Safe Drinking Water Act
- Enforcement and compliance assistance actions (formal and informal) and linkage data linking to violations.

2. SDWIS-FED's focus is on consolidating selected data, as previously described, on a national basis and making that limited set of data available to the states, regions and general public.

The SDWIS-FED database is a relational database, comprised of more than 50 tables. Many tables are look-up tables as well as association tables. Several tables, and attributes within existing tables, are not populated due to a decision to have a single data model for SDWIS-FED and SDWIS-STATE applications. Implementing this decision by re-engineering SDWIS-FED was determined to be infeasible and expensive. The resources to remove the unused structure

from the SDWIS-FED database were not expended. SDWIS-FED development optimized the storage of data, but not their access.

A data-warehousing model and associated on-line analytical processing capabilities now exist. Data is extracted and reformatted into the warehouse from the SDWIS-FED system quarterly update. A test data-warehousing environment will be operational shortly. It will likely provide Intranet access in the short term. Production Internet access is still to be defined.

Facility-specific inventory data is reported on an annual basis, while compliance data is reported quarterly with a one quarter reporting lag. The data is frozen each quarter after all processing and validation is completed.

3. The SDWIS-STATE system manages day-to-day PWSS program implementation.

SDWIS-STATE is software developed relatively recently to assist states without automated information management systems or the capability of developing one of their own. SDWIS-STATE operates in a client-server platform using a UNIX operating system or one of several versions of Windows operating systems. It uses an Oracle database for the back end, and the front end is written in C++. It contains 147 tables and 1,886 data elements. It addresses 726 analytes, 30 monitoring rules and 62 violation types.   EPA required reporting data is periodically extracted from SDWIS-STATE tables and converted to DTF format and submitted to SDWIS-FED where the data receives quality control checks and is then entered into SDWIS-FED tables.

4. SDWIS-FED data warehouse includes inventory (water system, water system facility, treatments, contacts, and location data) and compliance data; sample datasets, and results of PWS audits performed at states.

Periodically, the data warehouse extracts SDWIS-FED data into staging tables modeled after SDWIS-FED tables, and then performs additional quality checks on the data and transforms it, adding attributes and de-normalizing the data, and organizing it by subject. Several data marts are also periodically updated which contain subsets of the data (in the form of multi-dimensional star-schema cubes) which facilitate making analysis tools in the form of OLAP cubes and pivot tables, as well as an array of standard reports.

Fact tables include current violations, violations organized to facilitate trend analysis, analytical results from static and active sample datasets and data verifications findings. Conformed dimensions (which are basically the same as EPA Registries) facilitate information integration as they can be used with different data tables.

A. Baseline Architecture – *Business Processes, Data Model(s),* **Applications**

Applications – SDWIS-FED/-STATE uses a variety of applications:
- SDWIS-STATE software tools
- SDWIS-FED warehouse software tools

Background of Applications:

Reporting Toolkits -
The reporting of public drinking water inventory and noncompliance information to SDWIS-FED is supported by a variety of individually-developed state data systems as well as a personal computer (PC)-based, EPA-developed data entry tool (DTFWriter), and a full-featured local database application (SDWIS-STATE) developed by EPA for use by Primacy Agencies. DTFWriter was developed using Clipper™. The system can run on any computer supporting PC/MS DOS version 3.0 or higher. *Note: DTF is the data transfer format that SDWIS-FED is able to read. It will be replaced by extensible mark-up language (XML) in the near future.* Due to complexity and pending obsolescence of DTFWriter, EPA decided to develop Actions DTF, a short-term, stand-alone, single-purpose, PC-based application that supports Violations and Enforcements data entry. Actions DTF is a Microsoft™ (MS) Access® Windows application installed on a PC at the user site. Actions DTF was developed to assist state and regional PC users in the creation of a data file containing violations or enforcement actions information that can be input to the SDWIS-FED System. The software creates records in DTF that are required for entry of data into SDWIS-FED. States submit DTF files to the SDWIS-FED national database on a quarterly basis from the primacy agencies (states and EPA regions) that have been delegated PWSS oversight responsibility by EPA.

Data Entry subsystem –
This batch software (CLIST, JCL, CoolGen, COBOL, SAS, Assembler) performs input data editing and validation, constructs "update-in-place" transactions, posts data to the SDWIS-FED database, identifies, aggregates and creates error reports, and provides detailed and high level summaries of update status. Users are required to post the data to the EPA mainframe, and communicate data processing instructions to SDWIS-FED production control staff.

Data Retrieval subsystem –
This is the software that creates the user interface for canned retrievals of data from the SDWIS-FED database. There are over 15 standard reports designed for interactive batch access; storage of reports online or printed on high speed printers; or, provides access to the Platinum Report Facility, an *ad hoc* data retrieval tool.

SNC/Exception Tracking System –
This software provides support to EPA's enforcement and enforcement oversight programs via generated significant non-compliers (SNC) and exception records, three standard reports, and an on-line system for evaluation of noncompliance and enforcement data to allow regional modifications of the standard reports.

On-line Data Dictionary –
This MS-ACCESS application provides the data dictionary for the database.

Error Code Database –
This MS-ACCESS application provides a look-up for users debugging error reports to assist in understanding the nature of data entry errors and actions that need to be taken to correct those errors.

SDWIS-FED Data Warehouse –
EPA staffs operate, update and maintain a local data warehouse for data distribution and reorganization to enable easier access to SDWIS-FED data. Extract-transform-load (ETL) tools and procedures are utilized to extract data, transform it, and post it to the warehouse.
There are two ways of accessing drinking water violations and inventory data:
- Through the mainframe—standard reports or *ad hoc* queries using the platinum reporting facility (PRF), as well as through use of the Oracle Transparent Gateway (OTG), and
- Through the warehouse

MS-ACCESS custom queries provide access to the warehouse tables and many of the data marts, and through several pivot tables, which can be downloaded off the web.

Numerous analysis tools in the form of pivot tables and on-line analytical processing (OLAP) cubes have also been built and are continually refined. These include:
- GPRA, violations, and inventory analysis tools for trends analysis.
- Current violations and inventory (including contacts, locational data, treatments, etc.).
- An array of data quality analysis tools based on both data verifications and SDWIS-FED data that assess data quality, completeness and accuracy of violations data, percent correct compliance determination, rule implementation, timeliness of violations reporting, completeness of various required inventory elements including Surface Water Treatment Rule reporting and locational data.
- Several sample analysis tools for Unregulated Contaminant Monitoring Rule, 6-year Review, and Rounds 1 and 2 datasets.

<div style="border:1px solid">

A. Baseline Architecture – *Business Processes, Data Model(s), Applications and*
Technology

</div>

Technology – OGWDW data systems and warehouse operating environments:
- The SDWIS-FED Reporting System is designed to operate on the IBM mainframe computer system; the data is held in an IBM Database2 (DB2) database.
- SDWIS-STATE uses client-server architecture
- The SDWIS-FED data warehouse is in an SQL Server database.

1. The SDWIS-FED Reporting System

The SDWIS-FED operating environment incorporates use of the following software:
- IBM's Interactive System Productivity Facility (SIPF).

- IBM's DATABASE 2™ (DB2) Relational Database Management System (RDBMS).
- Platinum Technology's Platinum Report Facility™ (PRF).
- User dialogue screens implemented using IBM's Dialog Management Services (DMS).
- Control processing via IBM's Time Sharing Option (TSO) Command Lists (CLISTs).
- Report production performed through a combination of original COBOL programs and COBOL programs modified to utilize Structured Query Language (SQL) formulated from user-supplied selection criteria.

EPA headquarters staffs access the IBM mainframe via TCP/IP (Internet Protocol) -based communications between desktop devices and servers. EPA's ten regional office staffs and state primacy agency staffs access the mainframe system through the Internet using *IBM WebSphere Host On-Demand*™.

2. SDWIS-STATE

SDWIS-STATE uses client-server architecture and supports Oracle, MS SQL Server and IBM's DB2 database system as well as several operating systems including UNIX, WindowsNT, Windows 98 and Novell. The servers are housed at the state primacy agency offices and EPA provides SDWIS-STATE software.

3. SDWIS-FED Data Warehouse

The data warehouse is in an SQL Server database. The ETL tool is Microsoft Data Transformation Services (DTS). Several multidimensional OLAP cubes using MS Analysis Services software are available.

Data access for *ad hoc* queries is accomplished through MS-ACCESS databases, which have links to both SQL Server data warehouse tables, data marts, and SDWIS-FED mainframe tables. The Oracle Transparent Gateway is the means to access those tables that have not yet been pulled into the warehouse.

B. Target Architecture – Business Processes

Business Processes – OGWDW's Public Water System Supervision Program:
- Data flow will consist of end-to-end (from laboratories to EPA) electronic reporting that will utilize an eXtensible mark-up language (XML) format;
- Data will be exchanged securely using state and Agency e-commerce nodes; and
- Data flow will be received by a modern relational database (ORACLE), and will be accessed on line using SDWIS-FED data warehouse tools.

Background of Business Processes

The EPA Target Business Reference Model as presented in the document entitled "EPA Target Environmental and Health Protection Architecture," developed by the Office of Environmental Information, presents EPA's model for information integration. The Model for Information Integration (M4I)[1] was developed by the Information Integration Program and accepted by the Agency in July 2002. The M4I is a technical strategic framework that proposes an integration of data, applications and technology across the Agency. It consists of the following high-level functions:

- **Connect and Exchange** — Electronically connecting to transmit or access data
- **Process and Stage** — Data collection, cleanup, validation and approval for use
- **Store for Use** — Data storage, linkage and/or referencing for access and use
- **Use** — Data manipulation (potentially from multiple sources) to aid in learning, discovery and problem solving

Classifying major functions into these broad categories enables program and system managers across the Agency to think of information integration in general terms and to use common terms to discuss and plan for their programs' functional needs. Classifying systems by common functions helps identify areas where improvements to services as well as reductions in costs can be made by eliminating redundancies through the sharing of services. OGWDW intends to employ this high level classification of functions as it further refines its planning in support of system modernization that not only meets its immediate programmatic business needs but fully supports the enterprise business needs through conformance with the Agency's enunciated Enterprise Architecture.

Similarly, the shared business and program management support functions can be mapped to the OGWDW baseline business architecture as presented in Appendix B. The Agency's Enterprise Architecture Team is in the process of selecting tools that will enable the consistent mapping of program information into the Enterprise Architecture. A system called METIS may be the prescribed tool and adopted when it is fully supported by the Agency and available to the programs. It is OGWDW's intent to conform its target business architecture to the Agency's framework as outlined above.

While it is expected that many of OGWDW's business processes as described in the high level presentation in the baseline business process description will remain the same over the next three to five years, EPA will revise several business processes to meet the expanded information needs of OGWDW, which generally fall in the following categories:

- State oversight and assistance, including enforcement oversight
- National program oversight—key measures of program success, and program assessments
- Information to the public,
- Examining and implementing program improvements, including developing and evaluating regulations
- Support to program structure and function, including Homeland Protection, capacity development efforts

---

[1] *Model for Information Integration, A Preview of the Core Components of the EPA's Target Environmental Information Architecture (EIA)*, July 24, 2002.

- Conformance to the Agency's evolving Enterprise Architecture requirements including:
  - Participation in the Exchange Network, through adoption of XML as the data transfer language between states and EPA
  - Use of the System of Registries particularly the Facilities Registry, Chemical Registry, Biological Registry, proposed XML Registry, and the Metadata Registry
  - Use of the Central Data Exchange as OGWDW's data portal
  - Continued development of the OGWDW data warehouse and use of the Agency's data repository
  - Application of EPA data standards
  - Development of Trading Partner Agreements
- Integrate information
- Ways of integrating:
  - Geospatial tools
  - Data warehousing techniques
  - Agency Enterprise Architecture initiatives
    - ◆ Conformed dimensions/registries
    - ◆ Repositories

1. SDWIS-FED data warehouse plans through February 2004:

Format data warehouse: pull the data fields into the warehouse from SDWIS-FED that will be included in the future XML-formatted flows, warehouse them, and build an array of access tools from them. Information access should not change when the new data flow through CDX begins.

Internet access: work with Envirofacts to modernize and replicate standard reports using current off-the-shelf reporting tools that build reports from warehouse tables that OGWDW provides (i.e., the reports would be generated on line from the data warehouse tables rather than the mainframe). Integrate the warehouse tables and data marts into the central repository and registries, etc., using querying (ETL) tools when they're ready for us.

Intranet access: post warehouse tables and some data marts that supply pivot tables and standard reports on an intranet NT server on the EPA communication network (Tree). This will be used as an access server.

Move from the mainframe: move processing and warehouse storage from the mainframe to the NT server on the Agency network.

2. Next steps --

- Once new staging tables are built and modeled after the XML objects they will be populated from SDWIS-FED, replacing the current staging tables.
- Replicate and streamline the mainframe tools (Significant Non-complier Exception Tracking System or SETS) to run off the server or a PC.

- Explore the possibility of loading DTF directly into the warehouse, bypassing SDWIS-FED.
- With these steps taken, phase the mainframe out of the SDWIS data flow.

B. Target Architecture – Business Processes, **Data Model(s)**

Data Models for target architecture:
- SDWIS-FED: Transitioning to an architectural environment employing use of staging tables and other data warehousing technology will have a major impact on the current data model/structure of SDWIS-FED including the ultimate elimination of the system as it now exists. In the near term, the transference of data edits/verification to states and EPA regional offices could entail some structural changes as a consequence. Also, the replacement of DTF with XML will potentially involve structural changes to SDWIS-FED. There will likely be a period of operational overlap between SDWIS-FED and these new data structures until the new systems are fully functional in the operating environment.

- SDWIS-STATE: Transitioning from DTF to XML as the data exchange language between states and EPA and the data edit/validation responsibilities as described above will likely result in some structural modifications to the current SDWIS-STATE system to accommodate these changes. Fully Web enabling of SDWIS-STATE (beyond use of XML as the data exchange language) will entail other structural adaptations.

B. Target Architecture – Business Processes, Data Model(s), **Applications/Technology**

Applications
- By 12/2004, a new application called FedRep will be ready for distribution,
- FedRep is designed to run on local desktops and/or servers that will allow state and regional data providers (SDWIS-STATE) to validate the data, at their convenience and frequency, without the burden of moving the data to an EPA platform.

It will be designed to operate in environments where states have implemented SDWIS-STATE, or their own data management systems. It will take the EPA XML schema as input and use commercial off-the shelf (COTS) XML parser software for field and cross-field validations and preclude the need to develop custom software.

Technology
- Data will flow from states to EPA in XML format, the current industry standard.
- Draft schema still to be tested.

The drinking water draft schema was published in February 2003. The draft schema will be tested with data from volunteer states. The staging tables on the NT-Server will accept data from any state ready to exchange XML formatted data through CDX as soon as the schema is judged ready.

C. Transition Plan

Transition Plan
- Identify differences between baseline and target architectures;
- Phase out old components, phase in new; and,
- Share ISP document.

Step-by-Step Process:
OGWDW has been developing a transition or sequencing plan that provides a step-by-step process for moving from the baseline architecture to the target architecture and is the primary tool for management and investment decisions. OGWDW is or will be employing several techniques in its transition planning not the least of which is the sharing of this ISP document. Some of the techniques OGWDW is or will be employing include:

Gap Analysis –
The first step in transition planning is gap analysis – identifying the differences between the baseline and target architectures in all related architectural products. The "gap" between baseline and target architectures is overcome by a series of incremental builds that lead to the target environment.

Legacy, Migration and New Systems –
Legacy, migration and new systems make up the technical components for OGWDW's transition to the target environment. The key to prioritizing projects is the sequencing of the termination of systems, the phasing out of functionality and the timing of systems deployment, technology insertion, and the addition of new functionality. The sequencing plan addresses the dual operation of both the SDWIS-FED legacy system and the initial start-up of the new client-server based platform. The uninterrupted flow and management of data and its use by both legacy SDWIS and the new client-server system and its distribution is also addressed in the sequencing plan.

1. OGWDW decisions about sequenced investments are driven by high-level analyses about respective costs, benefits and risks as well as sequential technical and functional dependencies.

The OGWDW sequencing plan addresses the following migration issues:
- Continued operation of current system, e.g., SDWIS-FED.
- Existing system functionality being absorbed by another system, e.g., SDWIS-FED edits/validation routines moved into SDWIS-STATE.

- Legacy system transitions to migration and evolution into a new system, e.g., SDWS-FED to client-server to Web based system.
- Current system is planned for further evolution, e.g., OGWDW data warehouse to be expanded to include UIC and SWP, Sanitary Survey, and regulated occurrence data samples, as well.
- New systems developed during transition that become the permanent final system, e.g., PWSS existing data warehouse tools replace SDWIS-FED query tools as the main data access tool.

2. Planning the Migration

The rate of modernization, i.e., migration to the target architecture, is being planned in manageable increments to accommodate the organization's capacity to handle change. However, the final migration schedule will depend not only on the scope of the proposed changes but ultimately on priorities and availability of resources – budget, people and time constraints.

The interdependencies of systems among the projects and initiatives are the primary driving force in determining the sequence for implementing solutions. It is expected that the parallel development of Agency Enterprise Architecture requirements and possible associated changes in Agency planning, funding or budgeting processes, constitute significant unknowns and could greatly impact both development and implementation of the sequencing. A critical path analysis will be done based upon the estimated duration of each initiative and will be of great assistance in reducing the risk of project delay, cost overruns and adjusting to and minimizing the impacts of any new Agency requirements that may develop in process.

3. Fundamental Assumptions of the Transition Plan

Determining the steps involved in modernizing SDWIS, and establishing a timeframe for completing those steps, requires making some basic assumptions.
1. The budget for OGWDW SDWIS will remain at the same level over the transition period (through December 2004).
2. All existing SDWIS (FED and STATE) processes are maintained until the associated new replacement process is ready. The historical funding priority remains: reporting under new rules will be accommodated in federal data systems, especially in SDWIS/STATE.
3. There are no significant, unanticipated, new demands on the static budget.
4. Modernization will address three of the five identified historical data quality issues with SDWIS:
   - Difficult to get data into SDWIS,
   - High cost of processing and storing data in a mainframe environment, and
   - Difficult to get data out of SDWIS.
5. Modernization will *not* significantly address two of the five identified historical data quality issues with SDWIS:
   - Completeness of the data in SDWIS, and
   - Timeliness of the data submitted to SDWIS.

6. Additional, new data flows (e.g., occurrence data, sanitary survey data) that more directly support OGWDW programmatic decision-making, will be pursued simultaneously on a parallel track.

7. Any savings that accrue in the SDWIS budget due to modernization efforts will be applied to accommodating new rules (Ground Water Rule, Long Term 2 Enhanced Surface Water Treatment Rule, Stage 2 Disinfectants and Disinfection Byproducts Rule, etc.), and initiating information technology support to the Source Water Protection and Underground Injection Control Programs.

4. Transition Steps Already Completed:

The first step in transition planning, the gap analysis has been completed. The differences between the baseline and target architectures have been identified. The result of the analysis is a timeline of identified activities necessary to bridge the "gap" (differences) between the baseline and target architectures. The activities on the timeline will be accomplished using a series of incremental builds (based on annual budget levels) that will lead to the target environment.

Some aspects of the three data quality issues mentioned previously were immediately addressed since they were too critical to postpone.

- Difficulty in getting data into SDWIS:
  - Developed and distributed an On-line User's Guide,
  - Developed and distributed several error report tools,
  - MOU with the Office of Environmental Information (OEI) to flow Drinking Water data through the Agency Central Data Exchange (CDX) to:
    - ◆ further secure the current flow of DTF formatted data from states to EPA, and,
    - ◆ register states with CDX to initiate compliance with the new Agency Enterprise Architecture.
- High cost of processing and storing data in a mainframe environment:
  - Discontinue new development in legacy SDWIS-FED system during the transition period.
- Difficulty in getting data out of SDWIS:
  - Changed from DB2/COBOL programming languages that required using costly specialists to query the database to an on-line access approach using less expensive, off-the-shelf software (e.g., MS ACCESS) to provide pivot tables developed in a data warehouse environment for user-friendly interfaces with drinking water data.
  - Moved current OGWDW Data Warehouse from staff desktop computer to an NT Server.

5. Transition Steps: Ongoing and Future

Data In:
- Document (requirements, design), develop, implement, and train for XML schema:

- Replaces DTF as the format in which Drinking Water data is formatted for exchange from states to EPA, and
- Meets Agency recommendations/guidance.
- Document (requirements, design, User's Guide) develop, implement, and train for FedRep (validation software):
  - Increases Quality Assurance,
  - Puts data validation checks in the XML schema that will be loaded on state desktops to facilitate quality control by detecting data format problems at the point of data submission, and,
  - Facilitates data flow by minimizing data entry/format errors.
- Document (requirements, design, User's Guide), develop, implement, and train for Web-enabled version of the SDWIS-STATE software:
  - Allows access to software by anyone with browser software (rather than hard wire and/or compact disk),
  - Minimizes "version control" issues, and
  - Reduces need for costly/difficult to obtain/retain IT staff by states.

SDWIS0-FED Data Warehouse Processing and Access:
- Put less expensive (compared to mainframe) NT Servers (two of them) in EPA East basement for data warehouse processing and access:
  - Enhances report capability (e.g., Permanent Staging Tables for DTF, XML, Objects, etc.), and
  - Increases storage capacity; (e.g., occurrence data, sanitary survey data, etc.).

VI. CONCLUSION

- **OGWDW will implement the first stage of its Information Strategic Plan, addressing the public water system (PWS) facility and compliance data, by the end of 2004.**

At that time, the modernized components of the drinking water data flow will be able to accommodate end-to-end (from laboratories to EPA) electronic data transfers utilizing an eXtensible mark-up language (XML) format, where the data exchanges use secure state and Agency e-commerce nodes. This data flow will be received by a modern relational database (ORACLE), and will be accessible to the public and other stakeholders on line using data warehouse tools.

- **By the end of 2004, OGWDW will begin flowing occurrence (sample result) data for the first time.**

This data will be voluntarily provided by the states to support the Six-Year Review process.

The office will implement the second stage of this plan, including developing national data flows for Underground Injection Control compliance and Source Water Protection where the data will be available in the data warehouse, after 2004. When the warehouse is fully operational, EPA will have a comprehensive data set in one place. In the Safe Drinking Water Information System

(SDWIS) data warehouse, EPA will manage its ground water, drinking water, and source water protection data responsibilities under the Safe Drinking Water Act (SDWA). EPA will make these data publicly available as extensively as possible.

- **All of these efforts will comply with the Agency's Enterprise Architecture and Strategic Plan, and with federal government goals for well-managed, business-case-driven information technology investments.**

VII. MODERNIZATION TIMELINE

The following high-level timeline provides information on the anticipated scheduling of key SDWIS modernization tasks.